

Introduction

Eric E. Cahuzac

▶ **To cite this version:**

Eric E. Cahuzac. Introduction. Cahier des Techniques de l'INRA, 2010, Numéro spécial 2010, pp.3-5. hal-02654067

HAL Id: hal-02654067
https://hal.inrae.fr/hal-02654067

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Introduction

> *« Les sciences de l'homme et de la société sont avant tout des sciences de l'observation et l'expérimentation au sens strict n'est que rarement possible pour elles. »* (Silberman, 1999)[1].

L'accès aux bases de données en sciences humaines et sociales est crucial. Il permet aux chercheurs de tester les hypothèses des modèles théoriques et d'être en mesure de répondre de façon empirique aux questions que se pose la société.

Longtemps ces données ont été difficiles à collecter et demandaient la réalisation d'enquêtes parfois coûteuses, réalisées sur de petits échantillons. Les progrès de l'informatique et de la statistique ont peu à peu amélioré la collecte de données – effectuée principalement par les services de l'État ou les instituts de sondages et les chercheurs – et leur traitement pour offrir actuellement une information conséquente et de meilleure qualité dans des domaines très variés. Dans le même temps, la représentativité s'est accrue donnant accès à des niveaux d'analyses géographique et temporelle plus fins, faisant ainsi la part belle aux analyses micro économiques, aux modélisations longitudinales et spatialisées. Enfin, l'accès aux bases de données administratives s'est peu à peu assoupli sous l'impulsion des pouvoirs publics désireux d'exploiter cette grande quantité d'information à des fins d'expertise et de pilotage économique. Dernière en date, la directive INSPIRE[2] qui vise à favoriser la production et l'échange des données dans le domaine de l'environnement au sein de l'Union européenne permet de mettre à disposition de la recherche et de mutualiser dans un grand nombre de services de l'État un ensemble de données spatialisées très important.

Après avoir comblé en partie son retard – par rapport aux principaux pays producteurs de données – dans le domaine de la collecte, la France est en train d'améliorer notablement son processus de mise à disposition des données. Tout en finalisant cette dernière étape, c'est un nouveau défi qu'elle devra relever dans un futur proche, celui de la valorisation de ce gisement extraordinaire de micros données. Sans soigner la réalisation de cette dernière étape, le processus restera au stade de la construction d'entrepôts de données qui seront critiqués pour leur coût de mise en œuvre et de ce fait abandonnés. Dans le contexte actuel de rationalisation des moyens, cette étape de valorisation ne pourra s'appuyer comme les précédentes sur un seul acteur, les services de l'État gestionnaires des données. Elle doit en faire intervenir de nouveaux, et la recherche est ici le partenaire historique à privilégier pour répondre à la demande d'expertise sociale et valoriser scientifiquement ce gisement de connaissance que renferment les données.

[1] « Les sciences sociales et leurs données », Rapport à l'attention de Monsieur le ministre de l'Éducation nationale, de la recherche et de la technologie.
[2] Infrastructure for Spatial Information in the European Community. Directive 2007/2/CE du 14 mars 2007.

Nous nous situons donc actuellement au milieu d'un processus qui s'inscrit dans la durée. Il va nous amener à développer de nouvelles alliances entre acteurs publics, de nouvelles stratégies de recherches et de nouveaux outils d'acquisition et de diffusion de connaissances. Ces mutations posent bien évidemment de nouvelles questions, de nature juridique (propriété des données et des résultats, confidentialité, secret statistique, droit du citoyen,…) ou techniques (archivage et sauvegarde,…). Ces nouveaux questionnements qui sont souvent liés à la crainte d'une entrave à la vie privée du citoyen, doivent être abordés sereinement afin qu'ils ne représentent pas des freins au développement et à la diffusion de la connaissance.

En tant qu'acteur de la recherche agronomique, l'Inra a mis au centre de ces réflexions la question de la place des données dans ses recherches[3]. Dans les départements et les unités, les chercheurs ont mis en place des processus de gestion et d'archivage de ce patrimoine scientifique. Fait nouveau, des plateformes spécialisées dans le traitement et la valorisation des données voient peu à peu le jour et acquièrent souvent une renommée nationale, voire internationale.

Ce numéro spécial du *Cahier des Techniques de l'Inra* se propose de mettre en lumière différentes contributions collectives, témoignant de l'intérêt que portent les unités dans le traitement et la valorisation des données qu'elles gèrent. Il est dédié à des développements techniques s'appuyant sur des données en sciences humaines et sociales, plus précisément dans les disciplines de l'économie de la sociologie, de la géographie, jusqu'à l'écologie du paysage. Les domaines d'application seront variés mais ils répondent tous à la même logique : mettre à la disposition du chercheur, du citoyen ou d'autres institutions des outils permettant de structurer, de gérer et de valoriser de volumineuses bases de données. Enfin, la composante spatiale tient désormais une place importante dans la description des phénomènes et la plupart de ces réalisations y feront référence.

Nous présenterons tout d'abord deux exemples de réalisation de plateforme de traitements de données géo-localisées. La première, centrée sur la thématique du développement rural, est née d'un besoin non couvert de suivi et d'évaluation en France des politiques européennes liées à cette thématique. Partant de cet objectif, nous verrons comment l'Observatoire du développement rural (ODR) est devenu un outil coopératif mutualisant de nombreuses bases de données, au service de la recherche en sciences sociales et couvrant plus largement les questions de politiques agricoles, de développement rural et d'agro environnement. La seconde plateforme mutualise quant à elle des données acquises sur le terrain afin de traiter des questions relatives à la biodiversité. Nous verrons comment pour la plateforme LTSER[4] l'apport d'outils tels que les SIG[5] couplés avec des outils SGBD[6] a permis de mettre en réseau et de mutualiser des données de type écologiques, socio-techniques et environnementales.

La composante géographique des données est à nouveau mise à l'honneur dans la troisième et quatrième contribution de cette revue. Avec ODOMATRIX, on découvrira comment à partir de bases de données géographiques de l'IGN et de couches d'informations géographiques, la modélisation mathématique et la prise en compte des contraintes socio économiques permet de réaliser un distancier performant capable de fournir l'itinéraire le plus court en kilométrage ou en temps – pour les heures creuses et les heures de pointes – pour atteindre des pôles ou

[3] « Analyse de l'inventaire des bases de données scientifiques », rapport de mission, C. Christophe (2010).
[4] Long term sociological and ecological research
[5] Systèmes d'Information Géographique.
[6] Système de Gestion de Bases de Données.

des communes équipées en commerces et en services. La quatrième réalisation décrit d'un point de vue technique un phénomène économique complexe à modéliser qui est celui de la diversité des acteurs dans l'analyse de l'impact d'une politique. Nous nous placerons dans le cadre précis de l'impact des politiques publiques sur la dynamique des paysages. En alliant à un système multi-agent les capacités d'un SIG, nous appréhenderons la complexité du couplage de bases de données à la fois géographiques et hiérarchisées, sur une entité spatiale très fine : la parcelle.

Enfin, les deux dernières contributions illustrent dans deux domaines différents – celui de la comptabilité des exploitations et celui des échanges internationaux des industries agroalimentaires – la mise à disposition pour la recherche d'interfaces Web dédiées à l'interrogation et à la description de bases de données volumineuses. En effet, l'environnement Web est devenu aujourd'hui un environnement accessible à tous, alors que les logiciels d'archivage de sauvegarde et de traitements statistiques de données ont tous un environnement spécifique qu'il est difficile de maîtriser. La mise en forme des données dans un SGBD et la réalisation d'une interface développée avec des outils orientés pour le Web (PHP, MySQL) offre à tous des moyens d'interrogation et de traitement simples sur les bases de données sans pour autant posséder les compétences informatiques ou statistiques requises. Cette méthode favorise aussi une réelle politique de gestion des données dans une équipe, à savoir, la sauvegarde, la mise à jour et la documentation des bases de données, en implantant ces bases sur un serveur dédié.

Les articles que vous pourrez découvrir dans ce numéro spécial ont pour la plupart fait l'objet d'une présentation dans un atelier technique lors des deuxièmes journées de recherches en sciences sociales INRA - SFER[7]- CIRAD, en décembre 2008. Ils ont été écrits par des ingénieurs des départements SAE2[8] et SAD[9] qui contribuent dans leur équipe respective à l'amélioration de l'accessibilité aux bases de données pour la recherche.

Eric Cahuzac[10]
Chargé de Mission « Bases de Données »
Pour le département SAE2

[7] Société française d'économie rurale
[8] Sciences sociales agriculture et alimentation, espace et environnement
[9] Sciences pour l'action et le développement

[10] US0685 US ODR - Observatoire des programmes communautaires de développement rural – F-31326 Auzeville
☏ 05 61 28 53 41 ✉ EricCahuzac@toulouse.inra.fr